[BoldFont=Fira Sans SemiBold]Fira Sans Book Fira Mono

popquizbox[1] colback=nipun-lightblue!10, colframe=nipun-blue, boxrule=2pt, arc=3pt, left=8pt, right=8pt, top=8pt, bottom=8pt, title= **Quick Quiz 1**, fonttitle=**, coltitle=nipun-white, colbacktitle=nipun-blue, enhanced, attach boxed title to top left=xshift=0pt, yshift=-2pt, boxed title style=arc=3pt, boxrule=0pt**

definitionbox[1] colback=nipun-green!8, colframe=nipun-green, boxrule=1.5pt, arc=2pt, left=6pt, right=6pt, top=6pt, bottom=6pt, title= **Definition: 1**, fonttitle=**, coltitle=nipun-white, colbacktitle=nipun-green**

examplebox[1] colback=nipun-orange!8, colframe=nipun-orange, boxrule=1.5pt, arc=2pt, left=6pt, right=6pt, top=6pt, bottom=6pt, title= **Example: 1**, fonttitle=**, coltitle=nipun-white, colbacktitle=nipun-orange**

keypointsbox colback=nipun-blue!8, colframe=nipun-blue, boxrule=1.5pt, arc=2pt, left=6pt, right=6pt, top=6pt, bottom=6pt, title= **Key Points**, fonttitle=**, coltitle=nipun-white, colbacktitle=nipun-blue**

# Lagrangian and Duality

Nipun Batra

June 28, 2020

IIT Gandhinagar

Lectures heavily inspired by the Maths for Machine learning book

# Minimax Inequality

- Minimax inequality states:$\max_{\boldsymbol{y}} \min_{\boldsymbol{x}} q(\boldsymbol{x}, \boldsymbol{y}) \leqslant \min_{\boldsymbol{x}} \max_{\boldsymbol{y}} q(\boldsymbol{x}, \boldsymbol{y})$

# Minimax Inequality

- Minimax inequality states:$\max_{\boldsymbol{y}} \min_{\boldsymbol{x}} q(\boldsymbol{x}, \boldsymbol{y}) \leqslant \min_{\boldsymbol{x}} \max_{\boldsymbol{y}} q(\boldsymbol{x}, \boldsymbol{y})$
- We first prove For all $\boldsymbol{x}, \boldsymbol{y}$ $\quad \min_{\boldsymbol{x}} q(\boldsymbol{x}, \boldsymbol{y}) \leqslant \max_{\boldsymbol{y}} q(\boldsymbol{x}, \boldsymbol{y})$

# Minimax Inequality

- Let us choose $q(x, y) = xy$

# Minimax Inequality

- Let us choose $q(x, y) = xy$
- Let us first find $\max_{\boldsymbol{y}} q(\boldsymbol{x}, \boldsymbol{y})$

# Minimax Inequality

- Let us choose $q(x, y) = xy$
- Let us first find $\max_{\boldsymbol{y}} q(\boldsymbol{x}, \boldsymbol{y})$

# Minimax Inequality

- Let us choose $q(x, y) = xy$
- Let us first find $\max_{\boldsymbol{y}} q(\boldsymbol{x}, \boldsymbol{y})$

$$
\begin{array}{c|c|cccc}
x & 1 & 1 & 2 & 3 & 4 \\
\Downarrow & 2 & 2 & 4 & 6 & 8 \\
 & 3 & 3 & 6 & 9 & 12 \\
 & 4 & 4 & 8 & 12 & 16 \\
\hline
 & & 1 & 2 & 3 & 4 \\
 & & y & \Rightarrow & & 
\end{array}
$$

# Minimax Inequality

- For each value of $x$, we find $y$ that maximizes $q(x, y)$

| | | | | | |
|---|---|---|---|---|---|
| $x$ | 1 | 1 | 2 | 3 | 4 |
| $\Downarrow$ | 2 | 2 | 4 | 6 | 8 |
| | 3 | 3 | 6 | 9 | 12 |
| | 4 | 4 | 8 | 12 | 16 |
| | | 1 | 2 | 3 | 4 |
| | | $y$ | $\Rightarrow$ | | |

# Minimax Inequality

- For each value of $x$, we find $y$ that maximizes $q(x, y)$
- $y = 4$ maximizes $q(x, y)$ $\forall x$

| | | | | | |
|---|---|---|---|---|---|
| $x$ | 1 | 1 | 2 | 3 | 4 |
| $\Downarrow$ | 2 | 2 | 4 | 6 | 8 |
| | 3 | 3 | 6 | 9 | 12 |
| | 4 | 4 | 8 | 12 | 16 |
| | | 1 | 2 | 3 | 4 |
| | | $y$ | $\Rightarrow$ | | |

# Minimax Inequality

- For each value of $y$, we find $x$ that minimizes $q(x, y)$

| $x$ | | | | | |
|---|---|---|---|---|---|
| | 1 | 1 | 2 | 3 | 4 |
| $\Downarrow$ | 2 | 2 | 4 | 6 | 8 |
| | 3 | 3 | 6 | 9 | 12 |
| | 4 | 4 | 8 | 12 | 16 |
| | | 1 | 2 | 3 | 4 |
| | | $y$ | $\Rightarrow$ | | |

# Minimax Inequality

- For each value of $y$, we find $x$ that minimizes $q(x, y)$
- $x = 1$ minimizes $q(x, y)$ $\forall y$

| | | | | | |
|---|---|---|---|---|---|
| $x$ | 1 | 1 | 2 | 3 | 4 |
| $\Downarrow$ | 2 | 2 | 4 | 6 | 8 |
| | 3 | 3 | 6 | 9 | 12 |
| | 4 | 4 | 8 | 12 | 16 |
| | | 1 | 2 | 3 | 4 |
| | | $y$ | $\Rightarrow$ | | |

# Minimax Inequality

- We just showed For all $\boldsymbol{x}, \boldsymbol{y}$ $\quad \min_{\boldsymbol{x}} q(\boldsymbol{x}, \boldsymbol{y}) \leqslant \max_{\boldsymbol{y}} q(\boldsymbol{x}, \boldsymbol{y})$

| | | | | | |
|---|---|---|---|---|---|
| $x$ | 1 | 1 | 2 | 3 | 4 |
| $\Downarrow$ | 2 | 2 | 4 | 6 | 8 |
| | 3 | 3 | 6 | 9 | 12 |
| | 4 | 4 | 8 | 12 | 16 |
| | | 1 | 2 | 3 | 4 |

$y \quad \Rightarrow$

# Minimax Inequality

- We just showed For all $\boldsymbol{x}, \boldsymbol{y}$ $\quad \min_{\boldsymbol{x}} q(\boldsymbol{x}, \boldsymbol{y}) \leqslant \max_{\boldsymbol{y}} q(\boldsymbol{x}, \boldsymbol{y})$
- The equality occurs at $x = 1, y = 4$

| | | | | | |
|---|---|---|---|---|---|
| $x$ | 1 | 1 | 2 | 3 | 4 |
| $\Downarrow$ | 2 | 2 | 4 | 6 | 8 |
| | 3 | 3 | 6 | 9 | 12 |
| | 4 | 4 | 8 | 12 | 16 |
| | | 1 | 2 | 3 | 4 |
| | | $y$ | $\Rightarrow$ | | |

# Minimax Inequality

- Let us now find $\max_{\boldsymbol{y}} \min_{\boldsymbol{x}} q(\boldsymbol{x}, \boldsymbol{y})$

| $x$ | | | | | |
|---|---|---|---|---|---|
| | 1 | 1 | 2 | 3 | **4** |
| $\Downarrow$ | 2 | 2 | 4 | 6 | 8 |
| | 3 | 3 | 6 | 9 | 12 |
| | 4 | 4 | 8 | 12 | 16 |
| | | 1 | 2 | 3 | 4 |
| | | $y$ | $\Rightarrow$ | | |

## Minimax Inequality

- Similarly, let us now find $\min_{\boldsymbol{x}} \max_{\boldsymbol{y}} q(\boldsymbol{x}, \boldsymbol{y})$

| $x \Downarrow$ | | | | |
|---|---|---|---|---|
| 1 | 1 | 2 | 3 | **4** |
| 2 | 2 | 4 | 6 | 8 |
| 3 | 3 | 6 | 9 | 12 |
| 4 | 4 | 8 | 12 | 16 |
| | 1 | 2 | 3 | 4 |

$y \quad \Rightarrow$

# Minimax Inequality

- Similarly, let us now find $\min_{\boldsymbol{x}} \max_{\boldsymbol{y}} q(\boldsymbol{x}, \boldsymbol{y})$
- We can thus see our Minimax inequality $\max_{\boldsymbol{y}} \min_{\boldsymbol{x}} q(\boldsymbol{x}, \boldsymbol{y}) \leqslant \min_{\boldsymbol{x}} \max_{\boldsymbol{y}} q(\boldsymbol{x}, \boldsymbol{y})$

| | | | | | |
|---|---|---|---|---|---|
| $x$ | 1 | 1 | 2 | 3 | **4** |
| $\Downarrow$ | 2 | 2 | 4 | 6 | 8 |
| | 3 | 3 | 6 | 9 | 12 |
| | 4 | 4 | 8 | 12 | 16 |
| | | 1 | 2 | 3 | 4 |
| | | $y$ | $\Rightarrow$ | | |

# Revisiting the Lagrange multipliers

Our problem is of the form

$$\begin{array}{ccccc} \min_{\boldsymbol{x}} & f(\boldsymbol{x}) & & \\ \text{subject to} & g_i(\boldsymbol{x}) \leqslant 0 & \text{for all} & i = 1, \ldots, m \end{array}$$

# Revisiting the Lagrange multipliers

Our problem is of the form

$$\begin{array}{lll} \min_{\boldsymbol{x}} & f(\boldsymbol{x}) & \\ \text{subject to} & g_i(\boldsymbol{x}) \leqslant 0 & \text{for all} \quad i = 1, \ldots, m \end{array}$$

Idea: Convert constrained problem to an unconstrained problem

$$J(\boldsymbol{x}) = f(\boldsymbol{x}) + \sum_{i=1}^{m} \mathbf{1}\left(g_i(\boldsymbol{x})\right)$$

# Revisiting the Lagrange multipliers

Our problem is of the form

$$\begin{array}{lll} \min_{\boldsymbol{x}} & f(\boldsymbol{x}) & \\ \text{subject to} & g_i(\boldsymbol{x}) \leqslant 0 \quad \text{for all} & i = 1, \ldots, m \end{array}$$

Idea: Convert constrained problem to an unconstrained problem

$$J(\boldsymbol{x}) = f(\boldsymbol{x}) + \sum_{i=1}^{m} \mathbf{1}\left(g_i(\boldsymbol{x})\right)$$

where $1(z)$ is an infinite step function

$$\mathbf{1}(z) = \begin{cases} 0 & \text{if } z \leqslant 0 \\ \infty & \text{otherwise} \end{cases}$$

## Revisiting the Lagrange multipliers

Our problem is of the form

$$\begin{array}{lll} \min_{\boldsymbol{x}} & f(\boldsymbol{x}) & \\ \text{subject to} & g_i(\boldsymbol{x}) \leqslant 0 \quad \text{for all} & i = 1, \ldots, m \end{array}$$

Idea: Convert constrained problem to an unconstrained problem

$$J(\boldsymbol{x}) = f(\boldsymbol{x}) + \sum_{i=1}^{m} \mathbf{1}\left(g_i(\boldsymbol{x})\right)$$

where $1(z)$ is an infinite step function

$$\mathbf{1}(z) = \begin{cases} 0 & \text{if } z \leqslant 0 \\ \infty & \text{otherwise} \end{cases}$$

This would give infinte penalty if constraint is not satisfied. But, this formulation is hard to solve too.

# Revisiting the Lagrange multipliers

Idea: Introduce Lagrange multipliers ($\lambda_i \geq 0$) to "approximate" $J(x)$

$$\mathfrak{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i g_i(\boldsymbol{x})$$

# Revisiting the Lagrange multipliers

Idea: Introduce Lagrange multipliers ($\lambda_i \geq 0$) to "approximate" $J(x)$

$$\mathfrak{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i g_i(\boldsymbol{x})$$

What is the relationship between $\mathfrak{L}(\boldsymbol{x}, \boldsymbol{\lambda})$ and $J(\boldsymbol{x})$ given $\lambda_i \geq 0$?

# Revisiting the Lagrange multipliers

Idea: Introduce Lagrange multipliers ($\lambda_i \geq 0$) to "approximate" $J(x)$

$$\mathfrak{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i g_i(\boldsymbol{x})$$

What is the relationship between $\mathfrak{L}(\boldsymbol{x}, \boldsymbol{\lambda})$ and $J(\boldsymbol{x})$ given $\lambda_i \geq 0$?

When $\lambda \geqslant 0$, the Lagrangian $\mathcal{L}(x, \lambda)$ is a lower bound of $J(\boldsymbol{x})$. Hence, the maximum of $\mathfrak{L}(\boldsymbol{x}, \boldsymbol{\lambda})$ with respect to $\boldsymbol{\lambda}$ is

$$J(\boldsymbol{x}) = \max_{\boldsymbol{\lambda} \geqslant 0} \mathfrak{L}(\boldsymbol{x}, \boldsymbol{\lambda})$$

# Revisiting the Lagrange multipliers

$$J(\boldsymbol{x}) = \max_{\boldsymbol{\lambda} \geqslant 0} \mathfrak{L}(\boldsymbol{x}, \boldsymbol{\lambda})$$

# Revisiting the Lagrange multipliers

$$J(\boldsymbol{x}) = \max_{\boldsymbol{\lambda} \geqslant 0} \mathfrak{L}(\boldsymbol{x}, \boldsymbol{\lambda})$$

But, our original problem was minimizing $J(\boldsymbol{x})$, which is equivalent to:

$$\min_{\boldsymbol{x} \in \mathbb{R}^d} \max_{\boldsymbol{\lambda} \geqslant \mathbf{0}} \mathfrak{L}(\boldsymbol{x}, \boldsymbol{\lambda})$$

# Revisiting the Lagrange multipliers

$$J(\boldsymbol{x}) = \max_{\boldsymbol{\lambda} \geqslant 0} \mathfrak{L}(\boldsymbol{x}, \boldsymbol{\lambda})$$

But, our original problem was minimizing $J(\boldsymbol{x})$, which is equivalent to:

$$\min_{\boldsymbol{x} \in \mathbb{R}^d} \max_{\boldsymbol{\lambda} \geqslant \mathbf{0}} \mathfrak{L}(\boldsymbol{x}, \boldsymbol{\lambda})$$

Using the Minimax inequality, we can write:

# Revisiting the Lagrange multipliers

$$J(\boldsymbol{x}) = \max_{\boldsymbol{\lambda} \geqslant 0} \mathfrak{L}(\boldsymbol{x}, \boldsymbol{\lambda})$$

But, our original problem was minimizing $J(\boldsymbol{x})$, which is equivalent to:

$$\min_{\boldsymbol{x} \in \mathbb{R}^d} \max_{\boldsymbol{\lambda} \geqslant \mathbf{0}} \mathfrak{L}(\boldsymbol{x}, \boldsymbol{\lambda})$$

Using the Minimax inequality, we can write:

$$\min_{\boldsymbol{x} \in \mathbb{R}^d} \max_{\boldsymbol{\lambda} \geqslant \mathbf{0}} \mathfrak{L}(\boldsymbol{x}, \boldsymbol{\lambda}) \geqslant \max_{\boldsymbol{\lambda} \geqslant \mathbf{0}} \min_{\boldsymbol{x} \in \mathbb{R}^d} \mathfrak{L}(\boldsymbol{x}, \boldsymbol{\lambda})$$

# Revisiting the Lagrange multipliers

$$J(\boldsymbol{x}) = \max_{\boldsymbol{\lambda} \geqslant 0} \mathfrak{L}(\boldsymbol{x}, \boldsymbol{\lambda})$$

But, our original problem was minimizing $J(\boldsymbol{x})$, which is equivalent to:

$$\min_{\boldsymbol{x} \in \mathbb{R}^d} \max_{\boldsymbol{\lambda} \geqslant \mathbf{0}} \mathfrak{L}(\boldsymbol{x}, \boldsymbol{\lambda})$$

Using the Minimax inequality, we can write:

$$\min_{\boldsymbol{x} \in \mathbb{R}^d} \max_{\boldsymbol{\lambda} \geqslant \mathbf{0}} \mathfrak{L}(\boldsymbol{x}, \boldsymbol{\lambda}) \geqslant \max_{\boldsymbol{\lambda} \geqslant \mathbf{0}} \min_{\boldsymbol{x} \in \mathbb{R}^d} \mathfrak{L}(\boldsymbol{x}, \boldsymbol{\lambda})$$

We can write the dual objective as a function of $\lambda$ as $\mathfrak{D}(\boldsymbol{\lambda}) = \min_{\boldsymbol{x} \in \mathbb{R}^d} \mathfrak{L}(\boldsymbol{x}, \boldsymbol{\lambda})$

# Revisiting the Lagrange multipliers

- Primal objective:

$$\begin{array}{lll} \min_{\boldsymbol{x}} & f(\boldsymbol{x}) & \\ \text{subject to} & g_i(\boldsymbol{x}) \leqslant 0 \quad \text{for all} & i = 1, \dots, m \end{array}$$

# Revisiting the Lagrange multipliers

- Primal objective:

$$\begin{array}{lll} \min_{\boldsymbol{x}} & f(\boldsymbol{x}) & \\ \text{subject to} & g_i(\boldsymbol{x}) \leqslant 0 & \text{for all} \quad i = 1, \ldots, m \end{array}$$

- Or, primal objective $= J(x) \geq \max_{\boldsymbol{\lambda}} \mathfrak{D}(\boldsymbol{\lambda})$

# Revisiting the Lagrange multipliers

- Primal objective:

$$\begin{array}{ll} \min_{\boldsymbol{x}} & f(\boldsymbol{x}) \\ \text{subject to} & g_i(\boldsymbol{x}) \leqslant 0 \quad \text{for all} \quad i = 1, \ldots, m \end{array}$$

- Or, primal objective $= J(x) \geq \max_{\boldsymbol{\lambda}} \mathfrak{D}(\boldsymbol{\lambda})$
- Or, primal objective (in terms of x) $\geq$ dual objective (in terms of $\lambda$)

## Revisiting the Lagrange multipliers

- Primal objective:

$$\begin{array}{ll} \min_{\boldsymbol{x}} & f(\boldsymbol{x}) \\ \text{subject to} & g_i(\boldsymbol{x}) \leqslant 0 \quad \text{for all} \quad i = 1, \ldots, m \end{array}$$

- Or, primal objective $= J(x) \geq \max_{\boldsymbol{\lambda}} \mathfrak{D}(\boldsymbol{\lambda})$
- Or, primal objective (in terms of x) $\geq$ dual objective (in terms of $\lambda$)
- For SVM like formulations, primal objective is the same as dual objective (strong duality)

## Revisiting the Lagrange multipliers

- Primal objective:

$$\begin{array}{lllr} \min_{\boldsymbol{x}} & f(\boldsymbol{x}) & & \\ \text{subject to} & g_i(\boldsymbol{x}) \leqslant 0 & \text{for all} & i = 1, \ldots, m \end{array}$$

- Or, primal objective $= J(x) \geq \max_{\boldsymbol{\lambda}} \mathfrak{D}(\boldsymbol{\lambda})$
- Or, primal objective (in terms of x) $\geq$ dual objective (in terms of $\lambda$)
- For SVM like formulations, primal objective is the same as dual objective (strong duality)
- For some problems, there is a "daulity-gap" between the two objectives